Coherence and sufficient sampling densities for reconstruction in compressed sensing

Franz J. Király* Louis Theran†

February 13, 2013

Abstract

We introduce a concept called coherence for signals and constraints in compressed sensing. In our setting, we assume that the signal can be observed in finitely many features, and the set of possibly observable feature combinations forms an analytic variety which models the compression constraints. We study the question how many random measurements of the feature components suffice to identify all features. We show that the asymptotics of the sufficient number of measurements is determined by the coherence of the signal; furthermore, if the constraints are algebraic, we show that in general the asymptotics depend only on the coherence of the constraints, and not on the true signal, and derive results which explain the form of known bounds in compressed sensing. We exemplify our approach by deriving sufficient sampling densities for low-rank matrix completion and distance matrix completion which are independent of the true matrix.

1. Introduction

1.1. The general setting Compressed sensing can be formulated as the task of recovering a signal $x = (x_1, \ldots, x_n) \in \mathbb{K}^n$, where $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$, from noisy measurements of some of its entries x_i. Clearly this is impossible without some structural assumption; i.e., that x belongs to some subset $X \subseteq \mathbb{K}^n$ of signals with intrinsic properties that imply relations among the x_i. Well-known examples are: support on a sparse set of Fourier coefficients, the x_i are entries of a low-rank matrix, or that the x_i are squared distances. One can think of $\mathbb{K}^n$ as providing a parametric, or feature *representation* of x, and X as determining either *compression constraints* or *signal properties*.

A central question in compressed sensing is always which fraction of the observations are sufficient to reconstruct x with suitable accuracy. A common measurement model is the Erdős-Rényi sampling model[1], where each noisy entry x_i is observed independently with some probability p. Of particular interest is the asymptotic behavior. For a family of problems with $n \to \infty$, determine a function $f(n)$ such that $O(f(n))$ (random) measurements suffice for reconstruction of x. Most known results are of this kind, and many known results in different settings interestingly take the form

$$f(n) = c(x) \cdot \dim(X) \cdot \log^k(n) \tag{1}$$

*Machine Learning Group, TU-Berlin. franz.j.kiraly@tu-berlin.de

†Inst. Math., FU-Berlin. theran@math.fu-berlin.de

[1]the Erdős-Rényi model has, in most interesting cases, the same asymptotics as the uniform sampling model, where a fixed number of entries is observed uniformly

for some integer k, where $c(x)$ is some function measuring a property of x, often with $c(x)$ bounded independently of n, and $\dim(X)$ is the number of degrees of freedom of the signal space. For an overview on different compressed sensing problems and some asymptotic results, see for example Donoho (2006); Candès and Romberg (2007) or Candes and Wakin (2008).

A property of x which often plays a role in $c(x)$ is the *coherence*, or *incoherence* of x. In different setting, coherence of x is differently defined—classically for wave functions and time series, and, more recently, also for matrices Candès and Recht (2009)— but there is a common principle: The more incoherent a signal is, the more the signal itself is independent from the sampling process; as this also concerns information about the signal, the measurement of a more incoherent signal will provide more information than the measurement of a more coherent one. Thus, in the above formula, $c(x)$ will be smaller for incoherent x and bigger for coherent x.

In this paper, we provide an explanation for this behavior and a general framework for determining the measurement asymptotics of compressed sensing. Qualitatively, we will prove that

$$f(n) = \operatorname{coh}(X) \cdot n \log n \tag{2}$$

suffices for the (local) identifiability of a general and noiseless signal, under the condition that the constraint X is algebraic (or close to algebraic). Identifiability means that the set of signals is, in general, potentially reconstructible from the set of measurments. The quantity $\operatorname{coh}(X)$ which we will define fulfills $\operatorname{coh}(X) \leq 1$ and is called the coherence of (the constraints given by) X. Note that $f(n)$, as given above, does not depend anymore on properties of the signal x to reconstruct, only on properties of the constrained signal space X.

Moreover, if X is well-behaved, i.e., incoherent with the measurement process, then one will have $\operatorname{coh}(X) = O(\dim(X)/n)$, so in this case the asymptotics of the sufficient sampling density becomes

$$f(n) = \dim(X) \cdot \log n, \tag{3}$$

which also explains the form of the first equation, now with the influence of signal properties $c(x)$ removed.

1.2. Fixed coordinates and the logarithmic term Before continuing with specific examples, we want to point out an important conceptual point in the setting. Equations (2) and (3) involve a logarithmic term. Intuitively speaking, the number of measurements to obtain reconstruction is $\log n$ times at least the degrees of freedom $\dim(X)$ of the signal space. It can be counterintuitive that the $\log n$ term is in the bound, in particular in view of the following result, which is probably folklore and states that $\dim(X)$ measurements of general linear projection suffice.

Theorem 1.1. *Let $X \subseteq \mathbb{K}^n$ be an algebraic variety of dimension d, let $x \in X$. Let $\ell : \mathbb{K}^n \to \mathbb{K}^m$ be a generic linear map. If $m > d$, then x is uniquely determined by the values of $\ell(x)$, and the condition that $x \in X$.*

A proof is given in the appendix, see Theorem A.1. In view of this theorem—which is a statement on noise-free identifiability—one might now hypothesize that $f(n) = \dim(X)$, without the $\log n$ term, might always be the true sufficient asymptotics when there is no noise and x is sampled reasonably.

Nevertheless we claim that the bound with logarithm is the best possible one for the given setting, and that the $\log n$ term is not due to noise or sampling of x. Namely, lower bounds

including the $\log n$ term are known in the noise-free case which furthermore are independent of the particular sampling process of the signal, see Király et al. (2012a) for the case of matrix completion. The reason that the logarithmic term is necessary is that the coordinates x_i of the signal x are fixed, i.e., the bounds (2) and (3) hold for *any fixed* choice of coordinates, while Theorem 1.1 is a statement which holds for a *generic* choice of coordinates (i.e., with "random" coefficients) *Coordinate* projections. Since our coordinates are fixed and possibly degenerate, they are not generic in the sense of Theorem 1.1.

1.3. Examples for algebraic compressed sensing In this paper, we consider two algebraic compressed sensing problems - low-rank matrix completion and distance matrix completion - and derive sufficient reconstruction bounds by applying the coherence framework.

Low-rank matrix completion In the *low-rank matrix completion problem* the signal is an unknown matrix A, the measurements are the entries of A. The structural assumption is that A has rank at most r. The set of matrices of at most rank r is a *determinantal variety*, for example the variety of $(m \times n)$ matrices $X = \mathcal{M}(m \times n, r)$; it has $\dim(\mathcal{M}(m \times n, r)) = r(m + n - r)$ degrees of freedom. Results have been obtained in the case where the entries are observed with sampling probability p. Results of Candès and Tao (2010), Keshavan et al. (2010) and Király et al. (2012a) imply[2]:

Theorem 1.2. *Let $r \in \mathbb{N}$ be a fixed constant, and let A be an $(m \times n)$ matrix with $m \leq n$ that is* incoherent. *Then, there are constants $c > 0$ and $C > 0$ depending only on r such that, if each entry of A is sampled with probability p, then w.h.p.*[3]*, if*

$$p \leq c \cdot n^{-1} \log n,$$

then A cannot be reconstructed from the observed entries and the knowledge that $A \in X$, and if

$$p \geq C \cdot \lambda \cdot n^{-1} (\log n)^2$$

then A can be reconstructed from the observed entries and the knowledge that $A \in X$.

Due to a coupon collector's effect observed by Candès and Tao (2010), the order of p cannot be improved. We call attention to the fact that the assumption of *incoherence*, which will play a key role here as well, is on the true matrix A itself. Thus, if we take the view of Király et al. (2012a) which treats low-rank matrix completion as a parametric estimation problem, Theorem 1.2 bundles together: (a) the sampling of the true matrix A; and (b) the sampling of the observed entries.

Our coherence framework not only allows to generalize the statement but also to disentangle the sampling of the signal and the randomness in the observation process.

Distance matrix completion In the *distance matrix completion problem*, the signal is a distance matrix (also sometimes called similarity matrix), i.e., an $n \times n$ matrix A such that $A_{ij} = \|p_i - p_j\|^2$ if for some set of points $p_1, p_2, \ldots, p_n \in \mathbb{R}^r$. The structural assumption is that A is a distance

[2] What is stated here is a specialization, since the original theorems allow the rank to grow.

[3] w.h.p. = as $n \to \infty$, the probability of the statement which follows approaches 1

matrix, which is a non-trivial assumption: namely, the distance matrices make up a $(rn - \binom{r+1}{2})$-dimensional algebraic set $\mathcal{C}(r,n)$, a subset of $\binom{n+1}{2}$-space known as (real points of) the Cayley-Menger Variety (see Borcea (2002)). Again, entries are measured independently with fixed probability p. Coordinate projections of distance matrices are also known as *bar-joint frameworks*, which are central objects in *rigidity theory* (see, e.g., the monograph of Graver et al. (1993)); there, a framework is called *generically rigid* if A is reconstructible from the measurements up to finite choice. A well-studied question introduced by Thorpe (1983) relates to the rigidity properties of frameworks with the topology of an Erdős-Rényi random graph $G_{n,p}$. The only known (non-trivial) asymptotic results concern point dimensions $r \le 2$:

Theorem 1.3. *The following statements hold:*

1. *Let $A \in \mathcal{C}(2,n)$ be a distance matrix. Then, A can be reconstructed from the observed entries (i.e., $G_{n,p}$ is generically rigid) if and only if $p \ge n^{-1}(\log n + 2\log\log n + \omega(1))$ (Jackson et al., 2007).*
2. *The threshold for $G_{n,p}$ to contain any non-trivial rigid substructure is $p = c_2/n$, for a constant $c_2 \approx 3.588$; when a non-trivial rigid substructure emerges, it is linear-sized (Kasiviswanathan et al., 2011).*

The transition from rigid to flexible happens exactly at the threshold for the minimum degree of $G_{n,p}$ to reach 2, echoing the coupon collector's bound for low-rank matrix completion. The proofs of both parts of Theorem 1.3 are basically combinatorial, and rely, in an essential way on Laman's Theorem (1970). Asymptotic results for $r \ge 3$ are not known. Our framework permits the derivation of such results that generalize Theorem 1.3.

1.4. Contributions

Coherence and reconstruction In Section 2, we develop the theory of algebraic compressed sensing. Our main result elates the coherence of a a variety X to the sampling density of a generic (constrained) signal $x \in X$, which is needed to achieve reconstruction of x, up to finite choice. We show:

Theorem 1.4. *Let $X \subseteq \mathbb{K}^n$ be an irreducible algebraic variety, let Ω be the projection onto a set of coordinates, chosen independently with probability p, let $x \in X$ be a generic point on X. There is an absolute constant C such that if*

$$p \ge C \cdot \lambda \cdot \operatorname{coh}(X) \cdot \log n, \quad \textit{with } \lambda \ge 1,$$

then x is reconstructible from $\Omega(x)$ - i.e., $\Omega^{-1}(\Omega(x))$ is finite - with probability at least $1 - 3n^{-\lambda}$.

Here *generic* can be taken to mean that either x is sampled from a continuous probability density on X or taken from a certain Zariski dense subset of X. This generalizes Theorems 1.2 and 1.3 in several directions: (1) it applies more broadly, since it requires only a bound on the coherence of X; (2) the set of generic points in X is dense and includes some that are not "incoherent" in the sense of Theorem 1.2; (3) it doesn't rely on the specific combinatorial structure of the problem in the way that Theorem 1.3 does.

Low-rank matrices Our first application of Theorem 1.4 is to derive a variant of Theorem 1.2 which does not depend on properties of A.

Theorem 1.5. *Let $r \in \mathbb{N}$ be fixed, let A be a generic $(m \times n)$ matrix of rank at most r with $m \leq n$. Write $\bar{r} = \max(r, \log n)$* [4]*. There is a number C, depending only on r, such that if*

$$p \geq C \cdot \bar{r} \cdot n^{-1} \log n,$$

then w.h.p., A can be reconstructed from independent observations with probability p of its entries.

Moreover, if there is a Hadamard matrix of order n, the same conclusion holds with $p \geq C \cdot r \cdot n \log n$ and C now an absolute constant.

Proof. The first statement follows from Corollary 4.7, the second from Theorem 1.4 together with the coherence bounds from Propositions 2.5 and 4.3. □

We also show that the same bounds hold when A is taken to be symmetric.

Distance matrices Our second application is to distance matrices. The coherence of the Cayley-Menger variety is obtained by relating it to that of the determinantal variety. This is carried out in Section 5 using tools from Section 3. We obtain:

Theorem 1.6. *Let $r \in \mathbb{N}$ be fixed, let D be a generic distance matrix of points in r-space. There is a number C, depending only on r, such that if*

$$p \geq C \cdot n^{-1} (\log n)^2$$

then w.h.p., D can be reconstructed from independent observations with probability p of its entries.

Proof. This is an immediate consequence of Corollary 5.11. □

In the language of (Kasiviswanathan et al., 2011), Theorem 1.6 says that with the stated p, the random graph $G_{n,p}$ is generically rigid w.h.p. Because the minimum degree of a graph that is generically rigid in dimension d must be at least d, the order of the lower bound on p cannot be improved by more than a factor of $\log n$.

2. Coherence and signal reconstruction

In this section, we provide a framework for examining compressed sensing under algebraic constraints. First we introduce some formal concepts which describe the setting of compressed sensing under algebraic constraints, in particular the sampling process which we will assume to randomly and independently sample coordinate projections of the signal without repetition.

Definition 2.1. *Let $X \subseteq \mathbb{K}^n$ be an analytic variety. Fix coordinates $(X_1, \ldots, X_n)$ for $\mathbb{K}^n$. Let $S(p)$ be a the Bernouilli random experiment yielding a random subset of $\{X_1, \ldots, X_n\}$ where each X_i is contained in $S(p)$ independently with probability p. We will call the projection map $\Omega : X \to Y$ defined by $(x_1, \ldots, x_n) \mapsto (\ldots, x_i, \ldots \ : \ X_i \in S(p))$ of X onto the coordinates in $S(p)$, which is an analytic-map-valued random variable, a* random masking *of X with selection probability p.*

[4] Since r is fixed, for large enough k, $\bar{r} = \log n$.

The constraints defined by X play a crucial role in determining the sufficient sampling density which allows reconstruction of the signal. Namely, the central property of X which will determine the sufficient density is the so-called coherence, which describes the degree of randomness of a generic tangent flat to X; intuitively, it can be interpreted as the infinitesimal randomness of a signal.

Definition 2.2. *Let $H \subseteq \mathbb{K}^n$ be a k-flat[5]. Let $\mathcal{P} : \mathbb{K}^n \to H \subseteq \mathbb{K}^n$ the unitary projection operator onto H, let $e_1, \dots, e_n$ a fixed orthonormal basis of $\mathbb{K}^n$. Then the coherence of H with respect to the basis $e_1, \dots, e_n$ is defined as*

$$\operatorname{coh}(H) = \max_{1 \le i \le n} \|\mathcal{P}(e_i) - \mathcal{P}(0)\|^2.$$

When not stated otherwise, the basis e_i will be the canonical basis of the ambient space.

Remark 2.3. *Let $H \subseteq \mathbb{K}^n$ be a k-flat. Then the coherence $\operatorname{coh}(H)$ does not depend on whether we consider H as a k-flat in $\mathbb{K}^n$, or as a k-flat in $\mathbb{K}^m \supseteq \mathbb{K}^n$ for $m \ge n$ (assuming the chosen basis of $\mathbb{K}^m$ contains the basis of $\mathbb{K}^n$). Moreover, if $H \subseteq \mathbb{R}^n$, the coherence of H equals that of the complex closure of H.*

The coherence of a k-flat is bounded in both directions:

Proposition 2.4. *Let H be a k-flat in $\mathbb{K}^n$. Then, $\frac{k}{n} \le \operatorname{coh}(H) \le 1$, and the upper bound is tight.*

Proof. Without loss of generality, we can assume that $0 \in H$ and therefore that $\mathcal{P}$ is linear, since coherence, as defined in Definition 2.2, is invariant under translation of H.

First we show the upper bound. For that, note that for an unitary projection operator $\mathcal{P} : \mathbb{K}^n \to \mathbb{K}^n$ and any $x \in \mathbb{K}^n$, one has $\|\mathcal{P}(x)\| \le \|x\|$. Thus, by definition,

$$\operatorname{coh}(H) = \max_{1 \le i \le n} \|\mathcal{P}(e_i)\|^2 \le \max_{1 \le i \le n} \|e_i\|^2 = 1.$$

For tightness, take H as the span of $e_1, \dots, e_k$.

Let us now show the lower bound. We proceed by contradiction. Assume $\|P(e_i)\|^2 < \frac{k}{n}$ for all i. This would imply $k = n \cdot \frac{k}{n} > \sum_{i=1}^n \|\mathcal{P}(e_i)\|^2 = \|\mathcal{P}\|_F^2 = k$ which is a contradiction, where in the last equality we used the fact that orthonormal projections onto a k-dimensional space have Frobenius norm k. □

A similar definition of coherence, as in Definition 2.2, was used by Candès and Recht (2009), with different normalization. The following Proposition states how close the coherence of flats comes to the lower bound:

Proposition 2.5. *Denote $b(k, n) = \inf_{H \subseteq \mathbb{K}^n} \operatorname{coh}(H)$, where the infinum is taken over all k-flats H.*

(i) *Assume there exists an $(n \times n)$ Hadamard matrix[6]. Then $b(k, n) = \frac{k}{n}$.*

(ii) *There is a number C, depending only on k, such that $b(k, n) \le C \cdot \frac{\overline{k}_n}{n}$, where we write $\overline{k}_n = \max(k, \log n)$.*

[5]A k-flat is a linear subspace of dimension k which does not necessarily contain 0. Other names are affine subspace or affine linear variety.

[6]A Hadamard matrix is a not necessarily symmetric square matrix with entries ± 1 and mutually orthogonal rows.

Proof. (i) It suffices to provide a k-flat H such that $\mathrm{coh}(H) = \frac{k}{n}$ with respect to $e_1, \dots, e_n$. By applying a unitary transform, we can assume that $H = \mathrm{span}(e_1, \dots, e_k)$ and we need to provide a unitary system of coordinate vectors $v_1, \dots, v_n$ such that $\mathrm{coh}(H) = \frac{k}{n}$ with respect to $v_1, \dots, v_n$. We claim that we can take $(v_1, \dots, v_n) = \frac{1}{\sqrt{n}} M^\top$, where M is an $(n \times n)$ Hadamard matrix. Indeed, it holds that $\mathrm{coh}(H) = \sum_{i=1}^k (v_j^\top e_i)^2 = \sum_{i=1}^k \left(\pm\frac{1}{\sqrt{n}}\right)^2 = 1$.

(ii) follows from Lemma 2.2 of Candès and Recht (2009). □

Definition 2.6. *Let $X \subseteq \mathbb{K}^n$ be an (real or complex) irreducible analytic variety of dimension d (affine or projective). Let $x \in X$ a smooth point, and let $T_{X,x}$ be the tangent d-flat of X at x. We define*

$$\mathrm{coh}(x \in X) := \mathrm{coh}(T_{X,x}).$$

If it is clear from the context in which variety we consider x to be contained, we also write $\mathrm{coh}(x) = \mathrm{coh}(x \in X)$*. Furthermore, we define the coherence of X to be*

$$X = \inf_{x \in \mathrm{Sm}(X)} \mathrm{coh}(x),$$

where $\mathrm{Sm}(X)$ *denotes the smooth locus of X.*

Note that Remark 2.3 implies that the coherence $\mathrm{coh}(X)$ does not depend on the size of the ambient space. Also, if X is a k-flat, then the definitions of $\mathrm{coh}(X)$, given by Definitions 2.2 and 2.6 agree. These definitions, together with Proposition 2.5, imply:

Proposition 2.7. *Let $X \subseteq \mathbb{K}^n$ be an irreducible analytic variety. Then,* $\frac{1}{n}\dim X \leq \mathrm{coh}(X) \leq 1$.

Proof. Let $d = \dim X$. Irreducibility of X implies that, at each smooth point $x \in \mathrm{Sm}(X)$, the tangent space $T_{X,x}$ is a d-flat in $\mathbb{K}^n$. Both bounds then follow from Proposition 2.4. □

For ease of notation, we also define the incoherence as one minus coherence:

Definition 2.8. *For $X \subseteq \mathbb{K}^n$ an irreducible analytic variety, and $x \in X$, we define the incoherence*

$$\mathrm{incoh}(x) = 1 - \mathrm{coh}(x) \quad \textit{and} \quad \mathrm{incoh}(X) = 1 - \mathrm{coh}(X).$$

With these tools in place, we can prove our main result, which we recall from the introduction.

Theorem 1.4. *Let $X \subseteq \mathbb{K}^n$ be an irreducible algebraic variety, let Ω be the projection onto a set of coordinates, chosen independently with probability p, let $x \in X$ be a generic point on X. There is an absolute constant C such that if*

$$p \geq C \cdot \lambda \cdot \mathrm{coh}(X) \cdot \log n, \quad \textit{with } \lambda \geq 1,$$

then x is reconstructible from $\Omega(x)$ - i.e., $\Omega^{-1}(\Omega(x))$ is finite - with probability at least $1 - 3n^{-\lambda}$.

Sketch of proof: The argument, which is in Appendix B, integrates some ideas of Candès and Recht (2009) into our general algebraic setting.

Remark 2.9. *By the bounds given in Proposition 2.7, the best obtainable bound in Theorem 1.4 is $pn \geq C \cdot \lambda \cdot \dim(X) \cdot \log n$, with $\lambda \geq 1$.*

3. Coherence of subvarieties and secants

In the following, we will prove some bounds relating the coherence of different varieties to each other.

Lemma 3.1. *Let $H \subseteq \mathbb{K}^n$ be a k-flat, let $X \subseteq H$ be a subvariety. Then,* $\operatorname{coh}(X) \leq \operatorname{coh}(H)$.

Proof. We first prove the statement for the case where X is a flat; without loss of generality one can then assume that $0 \in X$. Let $\mathcal{P}'$ be the unitary projection onto X, similarly $\mathcal{P}$ the unitary projection onto H. Since $X \subseteq H$, it holds that $\|\mathcal{P}'x\| \leq \|\mathcal{P}x\|$ for any $x \in \mathbb{K}^n$. Thus, $\operatorname{coh}(X) \leq \operatorname{coh}(H)$.

The statement for the case where X is an irreducible variety follows from the statement for vector spaces. Namely, for $x \in X$, it implies $\operatorname{coh}(x \in X) \leq \operatorname{coh}(H)$, since the tangent space of X at x is contained in H. By taking the infimum, we obtain the statement. $\square$

Corollary 3.2. *As in Proposition 2.5, denote $b(k,n) = \inf_{H \subseteq \mathbb{K}^n} \operatorname{coh}(H)$, where the infinum is taken over all k-flats H.*

(i) *For any $\ell \in \mathbb{N}$, it holds that $b(k,n) \leq b(k+\ell, n+\ell)$.*

(ii) *If Payley's conjecture on Hadamard matrices[7] is true, then*

$$b(k,n) \leq \frac{k+c(n)}{n+c(n)}, \quad \text{where } c(n) = 4\left\lceil \frac{n}{4} \right\rceil - n.$$

Proof. (i) Let H be an generic $(k+\ell)$-flat in $\mathbb{K}^{n+\ell}$, let H_0 the n-flat spanned by the any n orthonormal basis vectors of $\mathbb{K}^{n+\ell}$. Then $H \cap H_0$ will be a generic k-flat in $H_0 \cong \mathbb{K}^n$, and Lemma 3.1 implies $\operatorname{coh}(H \cap H_0) \leq \operatorname{coh}(H)$. Since any k-flat in H_0 can be obtained in this way, and the possible H are dense in the set of all $(k+\ell)$-flats in $\mathbb{K}^{n+\ell}$, the statement follows. (ii) The statement follows from applying (i) to $\ell = c(n)$ and Payley's conjecture. $\square$

Note that Corollary 3.2 (i) is not a contradiction to the lower bound in Proposition 2.5, since it always holds that $\frac{k}{n} \leq \frac{k+\ell}{n+\ell}$ due to the fact that $k \leq n$.

Lemma 3.3. *Let $X, Y \subseteq \mathbb{K}^n$ be a analytic varieties, let $X + Y = \{\, x + y \;;\; x \in X, y \in Y \}$ be the sum of X and Y. Then,* $\operatorname{coh}(X) \leq \operatorname{coh}(X+Y)$.

Proof. Denote $Z = X + Y$, let $z \in Z$ be an arbitrary smooth point. By definition, there are smooth $x(z) \in X, y(z) \in Y$ such that $z = x(z) + y(z)$. Let T_z be the tangent space to $X + Y$ at z, let T_x be the tangent space of X at $x(z)$. An elementary calculation shows $T_x \subseteq T_z$, thus $\operatorname{coh}(x(z)) \leq \operatorname{coh}(z)$ by Lemma 3.1. Since z was arbitrary, we have $\operatorname{coh}(X) \leq \inf_{z \in \operatorname{Sm}(Z)} \operatorname{coh}(x(z)) \leq \operatorname{coh}(Z)$. $\square$

Remark 3.4. *In general, it is false that* $\operatorname{coh}(X+Y) \leq \operatorname{coh}(X) + \operatorname{coh}(Y)$. *Consider for example* $X = \operatorname{span}((1,1,1)^\top)$ *and* $Y = \operatorname{span}((1,-1,-1)^\top)$.

[7] Payley's conjecture states that there exists a $(n \times n)$ Hadamard matrix for every $4k, k \in \mathbb{N}$

4. Low-rank matrix completion: the determinantal variety

In this section, Theorem 1.4 is applied to obtain sampling densities for low-rank matrix completion in the symmetric and non-symmetric case.

Definition 4.1. *We will denote by $\mathcal{M}(m \times n, r)$ the set of $(m \times n)$ matrices in $\mathbb{K}$ of rank r or less, and by $\mathcal{M}_{sym}(n, r)$ the set of symmetric real resp. Hermitian complex $(n \times n)$ matrices of rank r or less, i.e.,*

$$\mathcal{M}(m \times n, r) = \left\{A \in \mathbb{K}^{m \times n}\ ;\ \mathrm{rk} A \leq r\right\}, \text{ and}$$
$$\mathcal{M}_{sym}(n, r) = \left\{A \in \mathbb{K}^{n \times n}\ ;\ \mathrm{rk} A \leq r, A^{\dagger} = A\right\}.$$

Since the matrices in $\mathcal{M}_{sym}(n, r)$ are symmetric resp. Hermitian, we will consider it as canonically embedded in $\frac{1}{2}n(n+1)$-space.

$\mathcal{M}(m \times n, r)$ is called the determinantal variety of $(m \times n)$-matrices of rank (at most) r, $\mathcal{M}_{sym}(n, r)$ the determinantal variety of symmetric $(n \times n)$-matrices of rank (at most) r.

As a corollary to Lemma 3.3, we obtain

Corollary 4.2. *For any m, n, r, it holds that*

$$\mathrm{coh}(\mathcal{M}(m \times n, r)) \leq \mathrm{coh}(\mathcal{M}(m \times n, r+1)) \quad \text{and} \quad \mathrm{coh}(\mathcal{M}_{sym}(n, r)) \leq \mathrm{coh}(\mathcal{M}_{sym}(n, r+1))$$

Proof. This follows from the equalities $\mathcal{M}(m \times n, r+1) = \mathcal{M}(m \times n, r) + \mathcal{M}(m \times n, 1)$ and $\mathcal{M}_{sym}(n, r+1) = \mathcal{M}_{sym}(n, r) + \mathcal{M}_{sym}(n, 1)$, and Lemma 3.3. □

The following main structural observation for determinantal varieties links the coherence of low-rank matrices to the coherence of the row and column spans.

Proposition 4.3. *Let $A \in \mathbb{K}^{m \times n}$, let H_m be the row span of A, and H_n the column span of A. Then,* $\mathrm{incoh}\,(A \in \mathcal{M}(m \times n, r)) = \mathrm{incoh}(H_n) \cdot \mathrm{incoh}(H_m)$ *and, if A is Hermetian,* $\mathrm{incoh}\left(A \in \mathcal{M}_{sym}(n, r)\right) = \mathrm{incoh}(H_n)^2 = \mathrm{incoh}(H_m)^2$.

Proof. The calculation leading to (Candès and Recht, 2009, Equation 4.9) shows in both cases that $\mathrm{coh}(A) = \mathrm{coh}(H_n) + \mathrm{coh}(H_m) - \mathrm{coh}(H_n)\,\mathrm{coh}(H_m)$, from which the statement follows. □

Corollary 4.4. *It holds that*

$$\mathrm{coh}(\mathcal{M}(m \times n, r)) = 1 - \sup_{H_m, H_n} \mathrm{incoh}(H_n)\,\mathrm{incoh}(H_m), \quad \text{and}$$
$$\mathrm{coh}(\mathcal{M}_{sym}(n, r)) = 1 - \sup_{H_n} \mathrm{incoh}(H_n)^2,$$

where the sup *range over all r-flats H_m in m-space and r-flats H_n in n-space. In particular, there are numbers C, C', depending only on r, such that*

$$\mathrm{coh}(\mathcal{M}(m \times n, r)) \leq C(mn)^{-1}(m\overline{r}_n + n\overline{r}_m - \overline{r}_m\overline{r}_n) \quad \text{and}$$
$$\mathrm{coh}(\mathcal{M}_{sym}(n, r)) \leq C' n^{-2}(2n\overline{r}_n - \overline{r}_n^2),$$

where we write $\overline{r}_k = \max(r, \log k)$.

Proof. The first statement follows from the fact that any pair of r-flats H_n and H_m in m-resp. n-space, there exists an $A \in \mathcal{M}(m \times n, r)$ such that the row resp. columns span of A is H_m resp. H_n, The second statement follows from the bound in Proposition 2.4. □

Remark 4.5. *Keep the notation of Proposition 2.5 and Corollary 3.2. If $b(r,m) = \frac{r}{m}$ and $b(r,n) = \frac{r}{n}$, then Corollary 4.4, together with Proposition 2.7 imply*

$$\operatorname{coh}(\mathcal{M}(m \times n, r)) = \frac{r}{mn} \cdot (m + n - r) = \frac{1}{mn} \dim(\mathcal{M}(m \times n, r)).$$

Corollary 4.6. $\operatorname{coh}(\mathcal{M}(n \times n, r)) = \operatorname{coh}(\mathcal{M}_{sym}(n, r))$.

Proof. $\operatorname{coh}(\mathcal{M}(n \times n, r)) \leq \operatorname{coh}(\mathcal{M}_{sym}(n, r))$ follows from Proposition 4.3 by considering $\mathcal{M}(n \times n, r) \subseteq \mathcal{M}_{sym}(n, r)$. For the converse, let $A \in \mathcal{M}(n \times n, r)$. It suffices to show that there is $M \in \mathcal{M}_{sym}(n, r)$ with $\operatorname{coh}(M) \leq \operatorname{coh}(A)$. Let H_1, H_2 be row and column span of A, such that $\operatorname{coh}(H_1) \leq \operatorname{coh}(H_2)$. Choosing an M with column (and thus also row) span H yields, by Proposition 4.3, an M with $\operatorname{coh}(M) \leq \operatorname{coh}(A)$. □

Corollary 4.7. *Let Ω be a random masking of $\mathcal{M}(m \times n, r)$, or $\mathcal{M}_{sym}(n, r)$ (in which case we set $m = n$), with sampling probability p. Let A be a generic matrix in $\mathcal{M}(m \times n, r)$, or $\mathcal{M}_{sym}(n, r)$. Write $\overline{r}_k = \max(r, \log k)$. Then there is an number C, depending only on r, such that if*

$$p \geq C \cdot \lambda \cdot (mn)^{-1} \cdot (m\overline{r}_n + n\overline{r}_m - \overline{r}_m \overline{r}_n) \log(mn), \quad \text{with } \lambda \geq 1,$$

then $\Omega^{-1}(\Omega(A))$ is finite with probability at least $1 - 3(mn)^{-\lambda}$.

5. Distance matrix completion: the Cayley-Menger variety

In this section, we will prove a bound on the coherence of the Cayley-Menger variety, i.e., the set of Euclidean distance matrices. The proof strategy is linking it to the case of symmetric low-rank matrices. We first introduce notation for the various occurring manifolds.

Definition 5.1. *Assume $r \leq m \leq n$. We will denote by $\mathcal{C}(n, r)$ the set of $(n \times n)$ real Euclidean distance matrices of points in r-space, i.e.,*

$$\mathcal{C}(n, r) = \left\{ D \in \mathbb{K}^{n \times n} \; ; \; D_{ij} = (x_i - x_j)^\top (x_i - x_j) \, \text{for some } x_1, \ldots, x_n \in \mathbb{K}^r \right\}.$$

Since the the elements of $\mathcal{C}(n, r)$ are symmetric, and have zero diagonals, we will consider $\mathcal{C}(n, r)$ as canonically embedded in $\binom{n}{2}$-space.

$\mathcal{C}(n, r)$ is called the Cayley-Menger variety of n points in r-space.

We will now continue with introducing maps related to the above sets:

Definition 5.2. *We define canonical surjections*

$$\begin{aligned} \varphi : (\mathbb{K}^r)^n &\to \mathcal{C}(n, r); & (x_1, \ldots, x_n) &\mapsto D \text{ s.t. } D_{ij} = (x_i - x_j)^\top (x_i - x_j), \\ \phi : (\mathbb{K}^r)^n &\to \mathcal{M}_{sym}(n, r); & (x_1, \ldots, x_n) &\mapsto A \text{ s.t. } A_{ij} = x_i^\top x_j. \end{aligned}$$

Note that φ, ϕ depend on r and n, but are not explicitly written as parameters in order to keep notation simple. Which map is referred to will be clear from the format of the argument.

We now define a "normalized version" of $\mathcal{M}_{sym}(n,r)$:

Definition 5.3. *Denote by* $\mathbb{S}^r = \{x \in \mathbb{K}^{r+1} \ ; \ x^\top x = 1\}$. *Then, define* $\mathcal{M}^\diamond_{sym}(n,r) := \phi\left((\mathbb{S}^r)^n\right)$. *Since* $\mathcal{M}^\diamond_{sym}(n,r)$ *contains only symmetric matrices with diagonal entries one, we will consider it as a subset of* $\binom{n}{2}$*-space.*

Remark 5.4. *The maps* φ, ϕ *are algebraic maps, and the sets* $\mathcal{C}(n,r), \mathcal{M}(m \times n, r), \mathcal{M}_{sym}(n,r), \mathcal{M}^\diamond_{sym}(n,r)$ *are irreducible algebraic varieties*[8].

Lemma 5.5. *For arbitrary* n, r, *one has* $\operatorname{coh}\left(\mathcal{M}_{sym}(n,r)\right) = \operatorname{coh}\left(\phi\left(\mathbb{K}^r\right)^n\right)$.

Proof. If $\mathbb{K} = \mathbb{C}$, then ϕ is surjective, so the statement follows. If $\mathbb{K} = \mathbb{R}$, note that the coherence of a general matrix does not depend on the variety it is considered in, since $\dim \mathcal{M}_{sym}(n,r) = \dim \phi\left(\mathbb{K}^r\right)^n$. Take $M \in \mathcal{M}_{sym}(n,r)$. Then, take any matrix $A \in \mathbb{R}^{n\times r}$ whose rows are a basis for the row span of M. Then, $AA^\top \in \phi\left(\mathbb{R}^r\right)$, and by Proposition 4.3, $\operatorname{coh}(M) = \operatorname{coh}(AA^\top)$. The statement follows from this. $\square$

The dimensions of the above varieties are classically known:

Proposition 5.6. *One has* $\dim \mathcal{C}(n,r) = \dim \mathcal{M}^\diamond_{sym}(n, r+1) = r \cdot n - \binom{r+1}{2}$, *and the dimensions are the same for the complex closures.*

Central in the proof will be the following map:

Definition 5.7. *For* $h \in \mathbb{K}$, *we will denote by*

$$\nu_h : \mathbb{K}^r \to \mathbb{S}^r \ ; \ x \to \frac{1}{\sqrt{x^\top x + h^2}}(x,h)$$

the map which considers a point $\mathbb{K}^r$ *as a point in the hyperplane* $\{(x,h) \ ; \ x \in \mathbb{R}^r\} \subseteq \mathbb{R}^{r+1}$ *and projects it onto* $\mathbb{S}^r$. *(if* $\mathbb{K} = \mathbb{C}$, *we fix any branch of the square root)*

Proposition 5.8. *For any* n, r, *it holds one has* $\operatorname{coh}(\mathcal{C}(n,r)) \le \operatorname{coh}(\mathcal{M}^\diamond_{sym}(n, r+1))$.

Proof. Lemma 5.10 implies that

$$\operatorname{coh}\left(\varphi\left((\mathbb{K}^r)^n\right)\right) \le \operatorname{coh}\left(\phi\left((\mathbb{K}^r)^n\right)\right),$$

the claim then follows from $\varphi\left((\mathbb{K}^r)^n\right) \subseteq \mathcal{C}(n,r)$ and Lemma 5.5. $\square$

We can bound the coherence of $\mathcal{M}^\diamond_{sym}(n,r)$ as follows:

Proposition 5.9. *There is a number* C, *depending only on* r, *such that* $\operatorname{coh}(\mathcal{M}^\diamond_{sym}(n,r)) \le C\frac{\overline{r}_n}{n}$, *where we write* $\overline{r}_n = \max(r, \log n)$.

[8] irreducibility for $\mathcal{C}(n,r), \mathcal{M}_{sym}(n,r), \mathcal{M}^\diamond_{sym}(n,r)$ follows from irreducibility of the respective ranges of the complex closure of φ, ϕ and surjectivity, irreducibility of $\mathcal{M}(m \times n, r)$ can be shown in a similar way; note that the real maps are in general not surjective

Proof. Lemma 2.2 of Candès and Recht (2009) proves that, for any fixed set of singular values, there exists a matrix $M \in \mathcal{M}(n \times n, r)$ with $\operatorname{coh}(M) \leq Cn^{-1}\overline{r}_n$ such that M has these singular values. By taking the singular values of M to be all one, and replacing M with a symmetric matrix M' having the same row or column span as M, as in the proof of Corollary 4.6, we see by Proposition 4.3 that $\operatorname{coh}\left(M' \in \mathcal{M}^{\diamond}_{sym}(n, r)\right) \leq \operatorname{coh}(M)$. $\square$

Our stated bounds on the number of samples required for distance matrix reconstruction then follow from the following lemma, which is proved in Appendix C.

Lemma 5.10. *Let $x_1, \ldots, x_n \in \mathbb{R}^r$. Let $D = \varphi(x_1, \ldots, x_n)$ and $A = \phi(\nu_h(x_1), \ldots, \nu_h(x_n))$, let T_D, T_A the respective tangent flats. Then, for $h \to \infty$, we have convergence $T_A \to T_D$, where we consider the tangent flats as points on the real Grassmann manifold of $\left(r \cdot n - \binom{r+1}{2}\right)$-flats in $\binom{n+1}{2}$-space.*

Corollary 5.11. *Let Ω be a random masking of $\mathcal{C}(n, r)$, with sampling probability p. Let D be a generic distance matrix. Then there is an number C, depending only on r, such that if*

$$p \geq Cn^{-1}(\log n)^2$$

then $\Omega^{-1}(\Omega(D))$ is finite with probability at least $1 - 3n^{-\lambda}$.

Proof. This follows from Theorem 1.4 and the coherence bounds from Propositions 5.8 and 5.9. $\square$

References

Ciprian S. Borcea. Point configurations and cayley-menger varieties. Preprint, arXiv:0207110, 2002. URL `http://arxiv.org/abs/math/0207110`.

Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009. ISSN 1615-3375. doi: 10.1007/s10208-009-9045-5. URL `http://dx.doi.org/10.1007/s10208-009-9045-5`.

Emmanuel J. Candès and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, 2007. ISSN 0266-5611. doi: 10.1088/0266-5611/23/3/008. URL `http://dx.doi.org/10.1088/0266-5611/23/3/008`.

Emmanuel J. Candès and Terence Tao. The power of convex relaxation: near-optimal matrix completion. *IEEE Trans. Inform. Theory*, 56(5):2053–2080, 2010. ISSN 0018-9448. doi: 10.1109/TIT.2010.2044061. URL `http://dx.doi.org/10.1109/TIT.2010.2044061`.

Emmanuel J. Candes and Mike B. Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21 –30, march 2008. ISSN 1053-5888. doi: 10.1109/MSP.2007.914731.

D.L. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289 –1306, april 2006. ISSN 0018-9448. doi: 10.1109/TIT.2006.871582.

Jack Graver, Brigitte Servatius, and Herman Servatius. *Combinatorial rigidity*, volume 2 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1993. ISBN 0-8218-3801-6.

Bill Jackson, Brigitte Servatius, and Herman Servatius. The 2-dimensional rigidity of certain families of graphs. *J. Graph Theory*, 54(2):154–166, 2007. ISSN 0364-9024. doi: 10.1002/jgt.20196. URL `http://dx.doi.org/10.1002/jgt.20196`.

Shiva Kasiviswanathan, Cristopher Moore, and Louis Theran. The rigidity transition in random graphs. In *Proc. of SODA'11*, 2011.

Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from a few entries. *IEEE Trans. Inform. Theory*, 56(6):2980–2998, 2010. ISSN 0018-9448. doi: 10.1109/TIT.2010.2046205. URL `http://dx.doi.org/10.1109/TIT.2010.2046205`.

Franz J. Király, Louis Theran, Ryota Tomioka, and Takeaki Uno. The algebraic combinatorial approach for low-rank matrix completion. Preprint, arXiv:1211.4116, 2012a. URL `http://arxiv.org/abs/1211.4116`.

Franz Johannes Király, Paul von Bünau, Frank Meinecke, Duncan Blythe, and Klaus-Robert Müller. Algebraic geometric comparison of probability distributions. *Journal of Machine Learning Research*, 13(Mar):855–903, 2012b. URL `http://jmlr.csail.mit.edu/papers/v13/kiraly12a.html`.

G. Laman. On graphs and rigidity of plane skeletal structures. *J. Engrg. Math.*, 4:331–340, 1970. ISSN 0022-0833.

David Mumford. *The Red Book of Varieties and Schemes*. Lecture Notes in Mathematics. Springer-Verlag Berlin Heidelberg, 1999. ISBN 054063293X.

M. Rudelson. Random vectors in the isotropic position. *J. Funct. Anal.*, 164(1):60–72, 1999. ISSN 0022-1236. doi: 10.1006/jfan.1998.3384. URL `http://dx.doi.org/10.1006/jfan.1998.3384`.

Walter Rudin. *Principles of mathematical analysis*. McGraw-Hill Book Co., New York, third edition, 1976. International Series in Pure and Applied Mathematics.

M. F. Thorpe. Continuous deformations in random networks. *J. Non-Cryst. Solids*, 57:355–370, 1983. URL `http://biophysics.asu.edu/thorpe/publications/84.pdf`.

A. Finiteness of Random Projections

The theorem, which will be proved in this section and which is probably folklore, states that for a *general* system of coordinates, a number of $\dim(X)$ observation is sufficient for identifiability.

Theorem A.1. *Let $X \subseteq \mathbb{K}^n$ be an algebraic variety or a compact analytic variety, let $\Omega : \mathbb{K}^n \to \mathbb{K}^m$ a generic linear map. Let $x \in X$ be a smooth point. Then, $X \cap \Omega^{-1}(\Omega(x))$ is finite if and only if $k \geq \dim(X)$, and $X \cap \Omega^{-1}(\Omega(x)) = \{x\}$ if $m > \dim(X)$.*

Proof. The theorem follows from the the more general height-theorem-like statement that

$$\operatorname{codim}(X \cap H) = \operatorname{codim}(X) + \operatorname{codim}(H) = \operatorname{codim}(X) + n - k,$$

where H is a generic k-flat, a proof of which can be found for example in the Appendix of (Király et al., 2012b). Then, the first statement about generic finiteness follows by taking a generic $y \in \Omega(X)$ and observing that $\Omega^{-1}(y) = H \cap X$ where H is generic if $k \leq \dim(X)$. That implies in particular that if $k = \dim(X)$, then the fiber $\Omega^{-1}(\Omega(x))$ for a generic $x \in X$ consists of finitely many points, which can be separated by an additional generic projection, thus the statement follows. $\square$

Theorem A.1 can be interpreted in two ways. On one hand, it means that any point on X can be reconstructed from exactly $\dim(X)$ random linear projections. On the other hand, it means that if the chosen coordinate system in which X lives is random, then $\dim(X)$ measurements suffice for (finite) identifiability of the map - no more structural information is needed. In view of Theorem 1.4, this implies that the log-factor and the probabilistic phenomena in identifiability occur when the chosen coordinate system is degenerate with respect to the variety X in the sense that it is intrinsically aligned.

B. Analytic reconstruction bounds and concentration inequalities

This appendix collects some analytic criteria and bounds which are used in the proof of Theorem 1.4. The first lemma relates local injectivity to generic finiteness and contractivity of a linear map. It is related to (Candès and Recht, 2009, Corollary 4.3).

Lemma B.1. *Let $\varphi : X \to Y$ be a surjective map of complex algebraic varieties, let $x \in X$, and $y = \varphi(x)$ be smooth points of X resp. Y. Let*

$$\mathrm{d}\varphi : T_x X \to T_y Y$$

be the induced map of tangent spaces[9]. Then, the following are equivalent:

(i) *There is an complex open neighborhood $U \ni x$ such that the restriction $\varphi : U \to \varphi(U)$ is bijective.*

[9] $T_x X$ is the tangent plane of X at x, which is identified with a vector space of formal differentials where x is interpreted at 0. Similarly, $T_y Y$ is identified with the formal differentials around y. The linear map $\mathrm{d}\varphi$ is induced by considering $\varphi(x + \mathrm{d}v) = y + \mathrm{d}v'$ and setting $\mathrm{d}\varphi(\mathrm{d}v) = \mathrm{d}v'$; one checks that this is a linear map since x, y are smooth. Furthermore, $T_x X$ and $T_y Y$ can be endowed with the Euclidean norm and scalar product it inherits from the tangent planes. Thus, $\mathrm{d}\varphi$ is also a linear map of normed vector spaces which is always bounded and continuous, but not necessarily proper.

(ii) $\mathrm{d}\varphi$ *is bijective.*

(iii) *There exists an invertible linear map* $\theta : T_y Y \to T_x X$.

(iv) *There exists a linear map* $\theta : T_y Y \to T_x X$ *such that the linear map*

$$\theta \circ \mathrm{d}\varphi - \mathrm{id},$$

where id *is the identity operator, is contractive*[10].

If moreover X *is irreducible, then the following is also equivalent:*

(v) $\varphi^{-1}(y)$ *is finite for generic* $y \in Y$.

Proof. (ii) is equivalent to the fact that the matrix representing $\mathrm{d}\varphi$ is an invertible matrix. Thus, by the properties of the matrix inverse, (ii) is equivalent to (iii), and (ii) is equivalent to (i) by the constant rank theorem (e.g., 9.6 in Rudin (1976)).

By the upper semicontinuity theorem (I.8, Corollary 3 in Mumford (1999)), (i) is equivalent to (v) in the special case that X is irreducible.

(ii)⇒ (iv): Since $\mathrm{d}\varphi$ is bijective, there exists a linear inverse $\theta : T_y Y \to T_x X$ such that $\theta \circ \mathrm{d}\varphi = \mathrm{id}$. Thus

$$\theta \circ \mathrm{d}\varphi - \mathrm{id} = 0$$

which is by definition a contractive linear map.

(iv)⇒ (iii): We proceed by contradiction. Assume that no linear map $\theta : T_y Y \to T_x X$ is invertible. Since φ is surjective, $\mathrm{d}\varphi$ also is, which implies that for each θ, the linear map $\theta \circ \mathrm{d}\varphi$ is rank deficient. Thus, for every θ, there exists a non-zero $\alpha \in \operatorname{Ker} \theta$. By linearity and surjectivity of $\mathrm{d}\Omega$, there exists a non-zero $\beta \in T_x X$ with $\mathrm{d}\Omega(\beta) = \alpha$. Without loss of generality we can assume that $\|\beta\| = 1$, else we multiply α and β by the same constant factor. By construction,

$$\left\| [\theta \circ \mathrm{d}\varphi - \mathrm{id}](\beta) \right\| = \|\theta(\alpha) - \beta\| = \|\beta\| = 1,$$

so θ cannot be contractive. Since θ was arbitrary, this proves that (iv) cannot hold if (iii) does not hold, which is equivalent to the claim. □

The second lemma is a consequence of Rudelson's Lemma, see Rudelson (1999), for Bernoulli samples.

Lemma B.2. *Let* $y_1, \ldots, y_M$ *be vectors in* $\mathbb{R}^n$, *let* $\varepsilon_1, \ldots, \varepsilon_M$ *be i.i.d. Bernoulli variables, taking value* 1 *with probability* p *and* 0 *with probability* $(1-p)$. *Then,*

$$\mathbb{E}\left(\left\| 1 - \sum_{i=1}^{M} \left(\frac{\varepsilon_i}{p} \right) y_i \otimes y_i \right\| \right) \leq C \sqrt{\frac{\log n}{p}} \max_{1 \leq i \leq M} \|y_i\|$$

with an absolute constant C, *provided the right hand side is* 1 *or smaller.*

[10] A linear operator $\mathcal{A}$ is contractive if $\|\mathcal{A}(x)\| < 1$ for all x with $\|x\| < 1$.

Proof. The statement is exactly Theorem 3.1 in Candès and Romberg (2007), up to a renaming of variables, the proof can also be found there. It can also be directly obtained from Rudelson's original formulation in Rudelson (1999) by substituting $\frac{\varepsilon_i}{\sqrt{p}} y_i$ in the above formulation for y_i in Rudelson's formulation and upper bounding the right hand side in Rudelson's estimate. $\square$

Now we proceed to the proof of the main theorem:

Theorem 1.4. *Let $X \subseteq \mathbb{K}^n$ be an irreducible algebraic variety, let Ω be the projection onto a set of coordinates, chosen independently with probability p, let $x \in X$ be a generic point on X. There is an absolute constant C such that if*

$$p \geq C \cdot \lambda \cdot \operatorname{coh}(X) \cdot \log n, \quad \textit{with } \lambda \geq 1,$$

then x is reconstructible from $\Omega(x)$ - i.e., $\Omega^{-1}(\Omega(x))$ is finite - with probability at least $1 - 3n^{-\lambda}$.

Proof. It suffices to prove the theorem for $\mathbb{K} = \mathbb{C}$; for $\mathbb{K} = \mathbb{R}$, we obtain the statement by replacing X by its Zariski closure in $\mathbb{C}^n$ and using the assumption that $x \in \mathbb{R}^n$; observe that $\#\Omega^{-1}(\Omega(x)) \leq \#(\Omega^{-1}(\Omega(x)) \cap \mathbb{R}^n)$.

By the definition of coherence, for every $\delta > 0$, there exists an x such that X is smooth at x, and $\operatorname{coh}(x) \leq (1+\delta)\operatorname{coh}(X)$. Now let $y = \Omega(x)$, we can assume by possible changing x that $\Omega(X)$ is also smooth at y. Let T_y, T_x be the respective tangent spaces at y and x. Note that y is a point-valued discrete random variable, and T_y is a flat-valued random variable. By the equivalence of the statements (iv) and (v) in Lemma B.1, it suffices to show that the operator

$$P = p^{-1}\theta \circ \mathrm{d}\Omega - \mathrm{id}$$

is contractive, where θ is projection, from T_y onto T_x, with probability at least $1 - 3n^{-\lambda}$ under the assumptions on p. Let $Z = \|P\|$, and let $e_1, \dots, e_n$ be the orthonormal coordinate system for $\mathbb{C}^n$, and $\mathcal{P}$ the projection onto T_x. Then the projection $\theta \circ \mathrm{d}\Omega$ has, when we consider T_x to be embedded into $\mathbb{C}^n$, the matrix representation

$$\sum_{i=1}^{n} \varepsilon_i \cdot \mathcal{P}(e_i) \otimes \mathcal{P}(e_i),$$

where ε_i are independent Bernoulli random variables with probability p for 1 and $(1-p)$ for 0. Thus, in matrix representation,

$$P = \sum_{i=1}^{n} \left(\frac{\varepsilon_i}{p} - 1\right) \cdot \mathcal{P}(e_i) \otimes \mathcal{P}(e_i).$$

By Rudelson's Lemma B.2, it follows that

$$\mathbb{E}(Z) \leq C \sqrt{\frac{\log n}{p}} \max_i \|\mathcal{P}(e_i)\|$$

for an absolute constant C provided the right hand side is smaller than 1. The latter is true if and only if

$$p \geq C^{-2} \log n \max_i \|\mathcal{P}(e_i)\|^2.$$

Now let U be an open neighborhood of x such that $\operatorname{coh}(z) < (1+\delta)\operatorname{coh}(X)$ for all $z \in U$. Then, one can write

$$Z = \sup_{y_1, y_2 \in U'} \left\| \sum_{i=1}^{n} \left(\frac{\varepsilon_i}{p} - 1 \right) \cdot \langle y_1, \mathcal{P}(e_i) \rangle \langle y_2, \mathcal{P}(e_i) \rangle \right\|$$

with a countable subset $U' \subsetneq U$. By construction of U', one has

$$\left\| \left(\frac{\varepsilon_i}{p} - 1 \right) \cdot \langle y_1, \mathcal{P}(e_i) \rangle \langle y_2, \mathcal{P}(e_i) \rangle \right\| \leq p^{-1}(1+\delta)\operatorname{coh}(X).$$

Applying Talagrand's Inequality in the form (Candès and Recht, 2009, Theorem 9.1), one obtains

$$P(\|Z - \mathbb{E}(Z)\| > t) \leq 3 \exp\left(-\frac{t}{KB} \log\left(1 + \frac{t}{2}\right)\right)$$

with an absolute constant K and $B = p^{-1}(1+\delta)\operatorname{coh}(X)$. Since δ was arbitrary, it follows that

$$P(\|Z - \mathbb{E}(Z)\| > t) < 3 \exp\left(-\frac{p \cdot t}{K \operatorname{coh}(X)} \log\left(1 + \frac{t}{2}\right)\right).$$

Substituting $p = C \cdot \lambda' \cdot \operatorname{coh}(X) \cdot \log n$, and proceeding as in the proof of Theorem 4.2 in (Candès and Recht, 2009) (while changing absolute constants), one arrives at the statement. □

C. Tangent space of the Cayley-Menger variety

In this appendix, we prove:

Lemma C.1. *Let $x_1, \ldots, x_n \in \mathbb{R}^r$. Let $D = \varphi(x_1, \ldots, x_n)$ and $A = \phi(\nu_h(x_1), \ldots, \nu_h(x_n))$, let T_D, T_A the respective tangent flats. Then, for $h \to \infty$, we have convergence $T_A \to T_D$, where we consider the tangent flats as points on the real Grassmann manifold of $(r \cdot n - \binom{r+1}{2})$-flats in $\binom{n+1}{2}$-space.*

Proof. Note that

$$\begin{aligned} D_{ij} &= x_i^\top x_i - 2x_i^\top x_j + x_j^\top x_j, \\ A_{ij} &= \frac{x_i^\top x_j + h^2}{\sqrt{x_i^\top x_i + h^2}\sqrt{x_j^\top x_j + h^2}}, \end{aligned}$$

An explicit calculation shows:

$$\begin{aligned} \left(\frac{\partial D}{\partial x_k}\right)_{ij} &= 2(\delta_{ki} + \delta_{kj})(x_i - x_j) \\ \left(\frac{\partial A}{\partial x_k}\right)_{ij} &= -(\delta_{ki} + \delta_{kj}) \frac{x_i(x_i^\top x_j + h^2)\sqrt{\frac{x_j^\top x_j + h^2}{x_i^\top x_i + h^2}} - x_j\sqrt{x_i^\top x_i + h^2}\sqrt{x_j^\top x_j + h^2}}{\left(x_i^\top x_i + h^2\right)\left(x_j^\top x_j + h^2\right)}, \end{aligned}$$

where δ_{ij} is the usual Kronecker delta. Thus,

$$\lim_{h\to\infty} h^2 \left(\frac{\partial A}{\partial x_k}\right)_{ij} = -\frac{1}{2}\left(\frac{\partial D}{\partial x_k}\right)_{ij}$$

which implies that both T_A converges to T_D in the Grassmann manifold when taking the limit $h \to \infty$; the statement directly follows. □